基于位置社交网络的个性化兴趣点推荐 *

韩笑峰，牛保宁†，杨 茸

(太原理工大学 计算机科学与技术学院，太原 030024)

摘 要：兴趣点(point-of-interest，POI)推荐是基于位置的社交网络(location-based social networks，LBSN)中一项重要的服务。针对目前推荐算法存在的噪声数据影响推荐质量，用户个性化程度低的问题，提出了一种个性化联合推荐算法。提出了引入POI的位置因素去除不可能或可能性较小的POI，形成初步候选集；综合考虑POI的类别、流行度及用户的社会行为，增加用户个性化的程度，提高推荐结果的质量。在Foursquare真实签到数据集上的实验，证明了提出的联合推荐算法与目前先进的算法相比，准确率提高11%，召回率提高8%。

关键词：兴趣点推荐；位置信息；分类信息；流行度信息；社会信息；位置社交网络

中图分类号：TP391 doi: 10.3969/j.issn.1001-3695.2017.11.0739

Personalized point-of-interest recommendation in location-based social networks

Han Xiaofeng, Niu Baoning†, Yang Rong

(*School of Computer Science & Technology*, *Taiyuan University of Technology*, *Taiyuan* 030024, *China*)

Abstract: Point-of-interest (POI) recommendation is an important service in location-based social networks (LBSNs) . For the current recommendation algorithm exists the problems of the noise data affects the recommended quality and low level of user personalization. Motived by this, this paper proposed a personalized joint recommendation algorithm (JRA) . JRA initially utilized the locality of user activity area to early filter the POIs which are impossible or less likely to be a result. For the received preliminary candidate set, then it also considered consider category factor and the popularity factor of POI, and the social behavior of the user to further improve the user experience. The experiments on real Foursquare check-in dataset demonstrate that the JRA compared with the current advanced algorithm, the accuracy rate increased by 11%, recall rate increased by 8%.

Key words: POI recommendation; locality of POI; category of POI; popularity of POI; social of POI; location based social network

0 引言

随着移动定位技术的进步和兴趣点(point-of-interests，POI)的增加（如商场、餐厅、公园、景点等），基于位置的社交网络(location-based social networks，LBSN)吸引了越来越多的用户。典型的LBSN有Foursquare、Gowalla、街旁、大众点评等。这些LBSN网站为用户提供位置签到、位置评论、位置与社交好友分享等功能，积累了大量可用于用户行为分析和个性化兴趣点推荐的数据。兴趣点推荐关联用户和兴趣点，既可以让用户迅速发现满足偏好的兴趣点，又可以让兴趣点找准自身定位，吸引相关用户，实现两者的双赢。

协同过滤技术[2]是常用的推荐技术，大量算法[3~6]都是以协同过滤作为基础的，它的基本思想是推荐的对象应当是与用户喜爱的对象相似，或者是与用户兴趣相似的其他用户喜爱的对象。在目前兴趣点推荐的研究中，主要存在以下两点不足：

a)未能提出有效的过滤机制消除噪声数据。庞大的用户签到数据中不可避免地混杂许多噪声数据，过多的噪声数据会导致推荐质量的降低。若能提前将原始数据中不符合用户行为习惯的签到数据筛选，可以有效提高推荐质量，并减少计算量，如将远离用户生活圈的兴趣点过滤。

b)个性化程度较低。每个用户的需求是不同的，个性化程度代表对用户需求的探索程度，影响推荐结果的质量。协同过滤算法聚类相似用户行为，体现用户的偏好[3~6]，即用相似用户的偏好代替用户自身偏好。这样做着重于反映和用户兴趣相类似的群体的社会化个性，忽视了对象自身的属性。用户访问过的兴趣点属性是用户偏好和需求最直观的表现，若能在相似用户的偏好的基础上，选择对象合适的属性加以考虑，可以维系用户的历史偏好，提高推荐结果的个性化程度。

收稿日期：2017-11-14；**修回日期：**2017-12-27 **基金项目：**国家自然科学基金资助项目（61572345）；国家科技支撑计划资助项目（2015BAH37F01）

作者简介：韩笑峰（1991-），男，山西太原人，硕士研究生，主要研究方向为数据挖掘、推荐系统；牛保宁（1964-），男（通信作者），山西太原人，教授，博导，博士，主要研究方向为大数据、数据库系统的自主计算与性能管理（1046184718@qq.com）；杨茸（1987-），女，山西洪洞人，博士研究生，主要研究方向为空间查询.

针对问题 a）本文提出基于位置的过滤算法。根据用户签到的地理特征对目标进行过滤，筛选出符合用户日常活动范围的兴趣点，去除噪声数据的干扰，提高推荐的质量。针对问题 b）本文利用用户访问过的兴趣点的分类、流行度来增强推荐结果的个性化程度。这两种属性可以方便地从签到数据中获取，带有明显的用户偏好，可以维系用户自身的历史兴趣，把它们与协同过滤算法得到的相似用户偏好相融合，提高推荐结果的个性化程度。

1　相关工作

LBSN 签到数据中包含多维度信息，与兴趣点和用户属性相关，利用 LBSN 签到数据可以提高兴趣点推荐质量。兴趣点推荐技术主要有以下两类。

a)基于内容的推荐。通过提取用户特征和兴趣点特征构建推荐模型。Gao 等人[6]将兴趣点特征、用户兴趣和用户情感相结合，将这三种类型信息合并到一个统一框架中，建立了一个推荐模型。Bao 等人[7]结合从兴趣点特征中得到的个人偏好和从数据集中分析出的专家信息，对兴趣点评分。这些推荐算法以被推荐对象的内容特征为主，推断用户的偏好，仅考虑用户和兴趣点本身，没有考虑用户间以及兴趣点之间的各种联系。

b)基于协同过滤的推荐算法。大致可以分为基于模型的协同过滤算法和基于记忆的协同过滤算法两类[8]。

基于模型的推荐算法的核心是使用用户—地点评分矩阵构建预测模型。Liu 等人[12]提出一种改进型奇异值分解模型，对用户的签到矩阵进行特征提取，有助于解决矩阵稀疏性的问题；但分解后的矩阵仍需还原，这需要很大的计算量。曹玖新等人[13]设置元路径特征集，利用随机游走算法度量节点间的关联度，用监督学习方法获得特征权值推断签到概率；然而元路径在收集阶段需要遍历整个网络的不同类型节点所有可能的链接情况，计算代价十分高昂。

目前的研究更多的集中在基于记忆的协同过滤算法[9]，并在此基础上加入其他因素来提高推荐质量。一种思路是通过挖掘用户之间的社会因素来提高兴趣点推荐质量。Konstas 等人[9]利用潜在因素模型获得用户社会关系中的相似性，再无缝衔接到基于用户的协同过滤中。另一种思路是利用位置因素提高推荐质量。Ye 等人[10]提出兴趣点的分布符合幂律分布，综合考虑兴趣点的距离因素和用户的社会因素进行协同过滤。Yuan 等人[11]认为用户行为受时空因素的制约，利用兴趣点的空间距离和时间差估计幂律分布，衡量访问位置对新位置的影响。以上研究仅考虑将社会因素和位置因素引入协同过滤算法中，本文在这两者的基础上作出改进，将其中的位置因素设置为预处理条件，增加了兴趣点的流行度因素和类别因素，提出计算 POI 分类流行度的方法，能有效提高推荐质量。

本文提出一种个性化联合推荐算法，综合考虑类别因素、流行度因素、位置因素、社交好友因素和用户历史签到行为，提出基于位置的过滤算法减少噪声和干扰，以提高推荐结果的质量。利用历史访问的兴趣点特征，结合用户的社会关系、历史行为提高用户个性化程度。

2　个性化联合推荐算法

2.1　问题描述

LBSN 中的兴趣点推荐是通过分析用户历史签到数据，为用户推荐未访问过的兴趣点。LBSN 中包含用户集 $U=\{u_1,u_2,\ldots\ldots,u_m\}$、兴趣点集 $L=\{l_1,l_2,\ldots\ldots,l_n\}$ 及用户在兴趣点的签到记录集合三类数据 $T=\{t_1,t_2,\ldots\ldots,t_s\}$。图 1 描述了一个简单的基于位置的社交网络图 $G=\{U,L,T\}$。其中包含若干用户、若干 POI 及三类相关关系—用户之间的好友关系，兴趣点之间的关联关系以及用户与兴趣点之间的签到关系。签到记录蕴涵有用户和 POI 这两种实体的三种关系。分析签到记录可以发现这些关系，从而提高推荐质量。本文提出的个性化联合推荐算法为用户提供一个包含 TOP-N 未曾访问过的 POI 的推荐列表。若之后用户对这些 POI 进行访问，则认为推荐的结果符合用户的判断，是高质量的推荐。

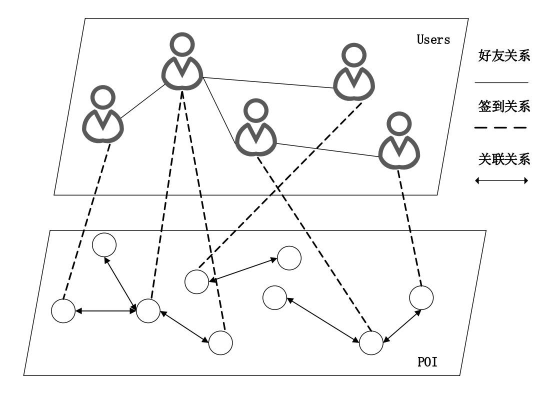


图 1　位置社交网络结构

为方便讨论，表 1 列出了本文中使用的一些重要符号。

表 1　符号

符号	描述
U	LBSN 中所有用户的集合(Users)
u_i	用户集中的一个用户：$u_i \in U$
L	LBSN 中所有 POIs 的集合(Locations)
l_j	POI 集合中的一个 POI：$l_j \in L$
C	LBSN 中所有分类的集合(Category)
c_k	类别集合的一个种类：$c_k \in C$
T	LBSN 中所有签到记录的集合(Tips)
t_g	签到集合的记录：$t_g\left(u_i,l_j\right)\in T$

2.2　兴趣点推荐中的位置因素

在实际的生活中，人们的活动往往局限于某一范围，反映在签到数据中，就是用户的签到行为发生在相对较小的地理空间内，称为签到的空间聚类现象[10]。地理学第一定律[14]指出：

任何事物都相关，只是相近的事物关联更紧密。因此，在对用户进行 POI 推荐时，位置因素是一个不可被忽视的因素，用户更趋向于访问距离较近的兴趣点。为了验证这个推断，本文做了下面的实验。

以 Foursquare 中真实数据集为例，对所有用户计算其访问过的任意两个 POI 之间的距离，并对得到的距离进行聚类，结果如图 2 所示。图中横坐标代表距离，纵坐标代表任意两个 POI 对间的平均距离小于横坐标指定距离区间的用户比例。例如，横坐标 5 km 对应的纵坐标代表平均距离处在 0~5 km 之间的用户比例。

图 2 显示，超过 89.3%以上的用户平均签到距离在 10 km 以内，可以认为当距离超过 10 km 时，用户访问该 POI 的可能性非常低。通过平均距离的计算，分析得出人们更倾向于访问与之前签到记录距离相近的兴趣点，且访问兴趣点的概率随着兴趣点距离的增加而逐渐降低。

因此，本文提出基于位置的过滤算法，对原始数据集根据距离信息进行过滤，将远离用户日常活动范围的 POI 排除，避免这些 POI 干扰推荐结果，提高推荐的质量。

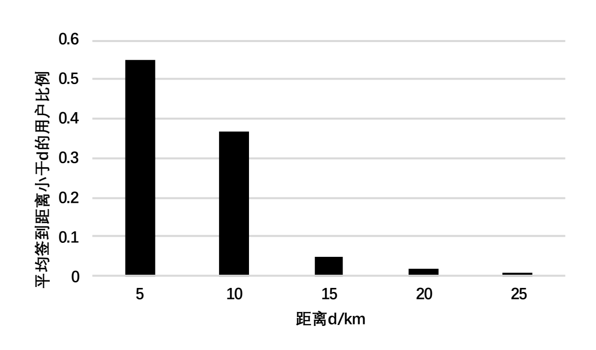


图 2　用户签到平均距离分布

基于上述分析，本文首先利用用户未访 POI 与已访问 POI 间的平均距离提出一种基于位置的过滤算法（location based filtering algorithm, LBFA），为用户推荐 TOP-N 个未访问的 POI。算法 1 列出了 LBFA 算法的伪代码，第 1~5 行首先扫描签到数据集的所有记录，选出所有属于指定用户 u_x 的签到兴趣点集合 LU。第 6 行生成用户仍未访问过的 POI 集合 Candlist。第 7~17 行依次遍历 Candlist 中每一个候选点，计算它与用户访问过的 POI 集合 LU 中兴趣点的平均距离（8~11 行），对得到候选点的平均距离进行判定，若其值小于规定的阈值(D_m=10 km)，则将这个点加入过滤后的最终候选点集合 CU 中（12~15 行），直到 Candlist 中所有点都判定完毕，返回过滤后的候选点集合 CU。

算法 1　基于位置的过滤算法

输入　用户 u_x、用户签到数据集 T、兴趣点集合 L。

输出　经过位置过滤的 POI 集合 CU。

1. for each $t_i \in T$ do
2. if($t_{i.u} = u_x$) then
3. 　$LU.add(t_i)$
4. end if
5. end for
6. $Candlist = L \setminus LU$
7. for each $l_i \in Candlist$ do
8. for each $l_j \in LU$ do
9. 　$D_i += Dist(l_i, l_j)$
10. end for
11. 　$D_i = D_i / |LU|$
12. if($D_i \le D_m$) then
13. 　$CU.add(l_i)$
14. end if
15. end for
16. return CU

2.3　兴趣点推荐中的分类流行度因素

LBSN 中的签到数据按照 POI 被分为不同的类别。类别信息隐含了 POI 的风格和提供的产品与服务。用户访问过的 POI 的类别信息可用于分析用户的个性化偏好。除了相似用户的偏好，类别信息也能体现用户的主观意愿。例如，当某用户在博物馆这个类别的 POI 签到记录数量远超其他类别时，可以认为该用户钟情于艺术收藏，当推荐 POI 时，应该优先推荐博物馆类别的 POI。

Foursquare 将所有的 POI 分为以下 8 大类：< Arts & Entertainment, College & University, Food, Great Outdoors, Buildings, Nightlife Spots, Shops, Travel Spots>，可以利用用户签到不同类别之间的数量关系来量化用户对不同类别的偏好程度。

如式(1)所示，首先统计用户每个类别的签到数量 $T(u,c) = \{t_n \in T \mid t_n.u = u \cap cat(t_n.l) = c\}$，再将其标准化为 0 到 1 之间的数值。式(1)中分子为用户访问某类别 POI 的签到数量，分母是用户访问过的所有类别中签到数量的最大值。通过式(1)每个用户都会得到一个对各个类别的偏好的得分向量，记为 $CAT(\mathrm{u}) = < cat(u,c_1),...,cat(u,c_8) >$，用于表示用户对于不同类别 POI 的偏好程度。

$$cat(u,c) = \frac{|T(u,c)|}{\arg_{c' \in C} MAX(|T(u,c')|)} \tag{1}$$

然而仅仅利用分类信息只能将用户偏好具体到类别，每个类别中又有许多的 POI，认为所有同类别 POI 对于用户是同等重要的显然是不合理的。为了得到同种类不同 POI 的权重，本文在类别的基础上引入了流行度因素。流行度即 POI 的受欢迎程度，可以反映 POI 所提供服务的质量。本文认为对于同类型的 POI，流行度越高，则 POI 的质量越高，推荐的优先级也应该越高。

从签到记录中可以得到以下两种标签：POI 总访客数量 $v(l_i)$ 和 POI 总签到数量 $t(l_i)$。POI 的访客数和签到数是同类别 POI 流行度最直观的表现，可以说明一个 POI 的受欢迎程度。由于 POI 之间的访客数和签到数可能相差很大，用式(2)计算已知类别的 POI 的流行度，采用调和平均数希望得到相对较大的

结果。

式(2)中 $CMAX(|t(l)|)=\arg_{l\in L,cat(l_i)=cat(l)} MAX(|t(l)|)$ 代表同类别的兴趣点中签到数量的最大值，$CMAX(|v(l)|)=\arg_{l\in L,cat(l_i)=cat(l)} MAX(|v(l)|)$ 代表同类别的 POI 中访客数量的最大值。

$$\mathrm{pop}(l_i)=\frac{2\times\frac{|t(l_i)|}{CMAX(|t(l)|)}\times\frac{|v(l_i)|}{CMAX(|v(l)|)}}{\frac{|t(l_i)|}{CMAX(|t(l)|)}+\frac{|v(l_i)|}{CMAX(|v(l)|)}} \tag{2}$$

综上，用户对候选 POI 的偏好得分 PS 利用式(3)来计算，其同时考虑了类别因素和流行度因素。

$$PS(u_i,l_j)=cat(u_i,c_k)\,pop(l_j) \tag{3}$$

2.4　兴趣点推荐中的社会因素

基于用户的协同过滤算法核心思想在于行为相似的用户应当有同样的偏好。传统的基于用户的协同过滤依据历史访问记录来寻找行为相似的用户。

在现实生活中，用户在购买不熟悉的物品前倾向于求助好友，类似地，在访问某个兴趣点时，比起陌生人或者 POI 供应商，用户更愿意相信自己的好友。同时，好友们也常常一起活动，例如好友会一起去看电影或者结伴去景点游玩，用户会去好友极力推荐的餐厅吃饭等。因此，好友之间常表现有共同的兴趣和相似的行为。研究证明，用户的所有首次访问记录中，超过 30%的 POI，其好友都曾经访问过[4]，因此有必要将用户的社会因素引入推荐系统，增加推荐的精度。

将社会因素引入基于用户的协同过滤算法之后，发生变化的量主要为用户访问候选 POI 的概率计算公式。

传统的基于用户的协同过滤算法计算用户 u_i 对任意候选 POI l_j 的访问概率计算公式为：$\hat{t}_{i,j}=\frac{\sum_U sim_{i,k}t_{k,j}}{\sum_U sim_{i,k}}$。其中：$sim_{i,k}$ 表示用户之间的相似度，可以有多种度量方法，如 Cosine、Jaccard 相似度以及皮尔逊相似度。其中 Cosine 相似度结果相对准确且方便计算[15]，因此选择这种方法计算用户相似度。

$sim_{i,k}=\frac{\sum_{l_j\in L}t_{i,j}t_{k,j}}{\sqrt{\sum_{l_j\in L}t^2_{i,j}}\sqrt{\sum_{l_j\in L}t^2_{k,j}}}$。$t_{i,j}$ 为签到记录表示用户 u_x 在兴趣点 l_j 的访问状态，若 $t_{i,j}=1$ 代表用户已在此处签到，$t_{i,j}=0$ 则代表用户还未在此签到。

引入社会因素后，用户 u_i 对任意候选 POI l_j 的访问概率计算公式为：$\hat{t}_{i,j}=\frac{\sum_{u_k\in F}SI_{i,k}\cdot t_{k,j}}{\sum_{u_k\in F}SI_{i,k}}$。其中：$F$ 代表用户 u_i 的好友集合；$SI_{i,k}$ 代表用户 u_k 对于用户 u_i 的社会影响因子。

用户的社会影响因子由相似程度和熟悉程度两部分组成。一方面，在实际的推荐过程中，并不是所有的好友都起正面的作用，有些用户尽管是社交好友，他们之间的兴趣相差极大。例如，本文一般会在社交网络中与长辈互相关注，但他们的偏好与本文相差极大。为了避免这种情况，好友之间的相似度必须考虑。另一方面，好友之间的推荐也不是同等重要的。一个点头之交建议的可信度与关系密切的好友的建议的可信度显然不一样。好友之间的熟悉程度也要考虑。因此使用式(4)表示用户的社会影响因子：

$$SI_{i,k}=sim_{i,k}\cdot fam_{i,k} \tag{4}$$

其中：$sim_{i,k}$ 表示好友间的相似度，仍使用 Cosine 相似度计算；$fam_{i,k}$ 表示好友间的熟悉程度，$fam_{i,k}=\frac{|F_i\cap F_k|}{|F_i\cup F_k|}$。这里由于是用户集合之间的计算，选择使用 Jarccard 相似度计算好友的熟悉程度。笔者认为用户间的共同好友数量越多，说明两者之间的关系越密切。

在实际计算时，将计算得到的所有候选 POI 的签到概率进行标准化得到用户关于 POI 的社会得分 SS，标准化公式为

$$SS(u_i,l_j)=\frac{|\hat{t}_{i,j}|}{\arg_{l_j\in L}MAX|\hat{t}_{i,j}|} \tag{5}$$

其中：$\hat{t}_{i,j}$ 代表候选 POI l_j 的签到概率；$\arg_{l_j\in L}MAX|\hat{t}_{i,j}|$ 代表所有候选点签到概率的最大值。

2.5　联合推荐算法

目前大多数流行的算法仅考虑将社会因素和位置因素引入协同过滤算法中，本文在这两者的基础上作出改进，将其中的位置因素设置为预处理条件，增加了兴趣点的流行度因素和类别因素，提出计算 POI 分类流行度的方法，与现有的协同过滤方法融合，最终形成了综合考虑 POI 的位置、类别、流行度、社会因素和用户行为的算法，即联合推荐算法(joint recommendation algorithm, JRA)。算法 2 列出了 JRA 算法的伪代码。第 1 行首先调用 LBFA 算法，对原始数据中的兴趣点集进行过滤，返回候选点集合 CU。第 2~6 行扫描候选点集中的所有兴趣点，分别调用 CompuCat( )和 CompuPop( )函数计算兴趣点的分类得分和流行度得分，将两者相乘得到兴趣点的偏好得分 PS。第 7~14 行调用 CompuSim( )和 CompuFam( )函数计算用户 u_x 和好友集合 F 中的每个用户 u_k 之间的相似度和熟悉度，将两者相乘得到用户的社会影响因子，把这个因子带入到好友协同过滤算法得到兴趣点的社会得分 SS。第 15 行根据参数 α (0< α <1)将 PS 和 SS 线性融合得到最终得分 S。第 16~18 行将候选点集 CU 根据得分 S 降序重新排序得到序列集 SU，选出 SU 中前 K 个兴趣点形成最终推荐结果集合 RS，完成推荐过程。

$$S(u_i,l_j)=\alpha\cdot SF(u_i,l_j)+(1-\alpha)\cdot SS(u_i,l_j) \tag{6}$$

算法 2　联合推荐算法

输入 用户 u_x、用户集合 U、用户签到数据集 T、兴趣点集合 L、兴趣点分类集合 C、用户 u_x 好友集合 F。

输出 TOP-N 个兴趣点组成的推荐列表 RS。

```
CU = LBFA(u_x, T, L)
for each l_j ∈ CU do
    cat(u_x, c_i) ← CompuCat(u_x, c_i)
    pop(l_j, c_i) ← CompuPop(l_j, c_i)
SF(u_x, l_j) = cat(u_x, c_i) · pop(l_j, c_i)
end for
for each l_j ∈ CU do
for each u_k ∈ F do
    sim(u_x, u_k) ← CompuSim(u_x, u_k)
     fam(u_x, u_k) ← CompuFam(u_x, u_k).
  SI(u_x, u_k) = sim(u_x, u_k) fam(u_x, u_k).
    SS(u_x, l_j) =FBCF(SI)
  end for
end for
S(u_x, l_j) = α · SF + (1 − α) · SS
SU = Reorder(CU, S(u_x, l_j))
RS = SU.top(k)
return RS
```

3 实验设计

3.1 数据集描述

本文采用典型的 LBSN 网站 Fousquare 的公开用户签到数据集。数据收集自美国的两个大型城市——纽约和洛杉矶。其中纽约数据集包括 49 062 个用户的 221 128 条签到数据，兴趣点的数量为 92 018 个。洛杉矶数据集包括 31 544 个用户的 104 478 条签到数据，兴趣点的数量为 70 241 个。将 Foursquare 的两个数据集中每个用户的签到数据按时间顺序划分，其中的 75%选为训练集，余下的 25%作为测试集。选取数据的相关信息如表 2 所示。

表 2　数据集结构

数据类型	包含主要属性
users	User id, Gender, City, Friend id
venues	Venue id, latitude longitude, check-in ,visit, category
tips	User id, Venue id, latitude longitude, time

3.2 评价指标

本文选取两个在推荐算法中应用最为广泛的评价指标：准确率 precision@N 和召回率 recall@N，分别如式(7)和(8)所示，N 代表最终推荐结果的数量。准确率是指算法推荐结果中用户实际访问的兴趣点数量占推荐结果总数的比例，反映推荐的准确性。召回率是指算法推荐结果里用户访问的兴趣点数量占用户实际访问兴趣点总数的比例，反映推荐的全面性。其中：*R(U)* 代表推荐算法在训练集执行后得到的兴趣点推荐列表；而 *T(u)* 代表用户在测试集上的实际签到的兴趣点列表。

$$\mathrm{Precision} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \tag{7}$$

$$\mathrm{Recall} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(U)|} \tag{8}$$

3.3 参数 α 选取

在第 2.5 节提到需要确定参数 α (0<α<1)的取值，调节用户的偏好得分 *PS* 和社会得分 *SS* 在推荐结果中所占的比例，当 α 的值越大时，通过用户的兴趣点特征得到的偏好得分对结果的影响比较大；反之，通过用户好友协同过滤得到的社会得分所占的比例较大，通过在实际数据集上的测试来确定 α 的取值。

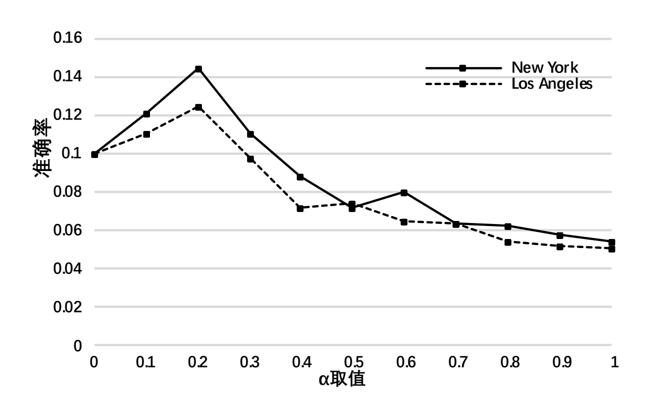


图 3　参数 α 的取值对应的准确率

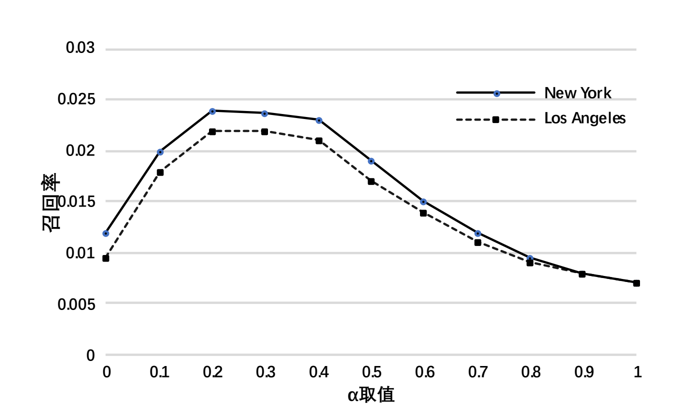


图 4　参数 α 的取值对应的召回率

图 3、4 分别是比较 α 在不同取值下对应的准确率和召回率变化趋势。当 α 取值为 0.2 左右时，可以同时获得最高的准确率和召回率。因此在之后的对比实验时，将 α 值默认设置为 0.2。

3.4 实验性能比较

为了验证本文提出的个性化联合推荐算法的性能，把它与两个基础推荐算法以及目前先进的推荐算法作比较，比较的算法如表 3 所示。

表 3 比较的推荐算法

算法(简称)	算法描述
User based CF(U) [2]	基于用户的历史签到数据计算用户之间的相似度，再根据相似用户记录计算候选兴趣点得分。
Friend based CF(F) [9]	基于用户历史签到数据和社会关系计算用户之间的相似度，再根据相似用户的访问记录推荐兴趣点。
USG(G)[10]	同时考虑用户的历史签到数据和兴趣点的社会关系和地理因素，将三者线性融合得到结果。
JRA(J)	本文提出的个性化联合推荐算法，综合考虑了分类因素、流行度因素、位置因素、社交好友因素和用户历史签到行为。

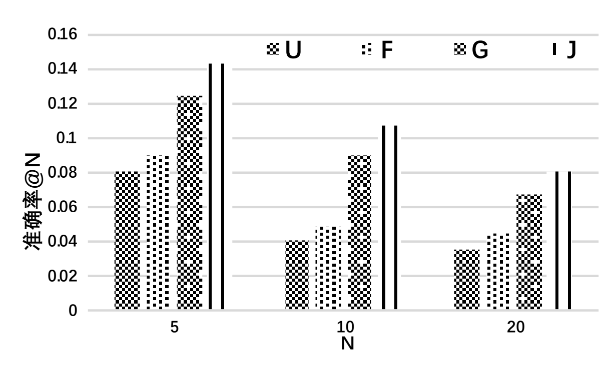


图 5 纽约数据集推荐结果的准确率

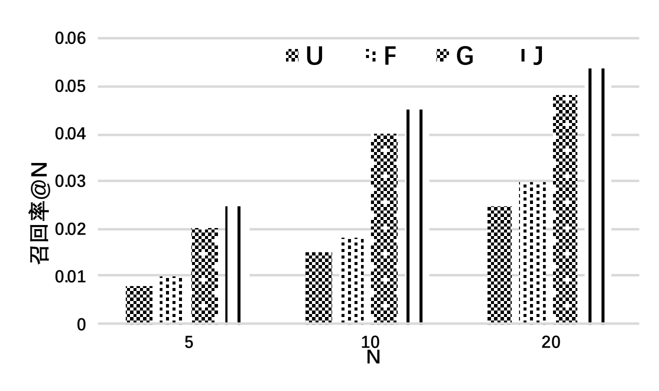


图 6 纽约数据集推荐结果的召回率

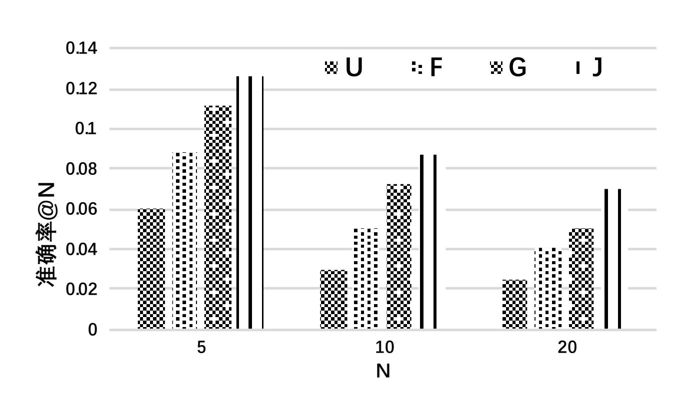


图 7 洛杉矶数据集推荐结果的准确率

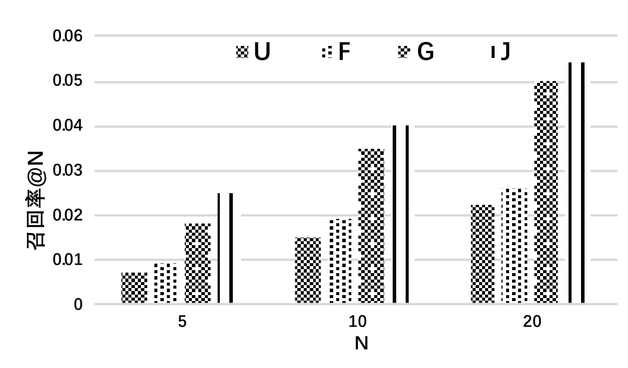


图 8 洛杉矶数据集推荐结果的召回率

针对两个数据集，将本文提出的个性化联合推荐算法与三个不同的推荐算法作了比较。其中图 5、6 对应 Foursquare 的纽约数据集，图 7、8 对应 Foursquare 的洛杉矶数据集，展示了各种算法 TOP-N（N=5, 10, 20）的推荐性能。由图 5~8 分析可得，算法 U 作为一个基础的协同过滤算法，结果的准确率和召回率都最低；算法 F 在 U 的基础上引入了社会因素，推荐结果略高于 U，这证明了社会因素在推荐过程中起了积极因素；算法 G 在算法 F 的基础上进一步引入了位置因素，得到相对较高的准确率和召回率，说明位置因素在兴趣点推荐时不可忽视，其可以很大程度上提高推荐质量；而本文提出的联合推荐算法 J 在两个数据集的准确率和召回率都高于其他几个算法，与其中表现较好的 G 算法相比平均提高了 11%的准确率和 8%的召回率，可见基于位置的预处理和引入类别因素、流行度因素可以显著提高推荐的质量。

3.5 实验开销比较

本实验中，查询时间定义为在所有 POI 上计算一个用户 u 的推荐分数的平均时间（查询时间不包括对查询结果中 POI 的排序时间）。不同算法的查询时间如表 4 所示。结果显示，传统的协同过滤算法 U 的查询时间最短，增加考虑因素会使查询时间适当增长，但不同的查询时间相差很小，相较于结果准确率和召回率的提升，这些时间开销可以接受。

表 4 不同算法的查询时间/s

推荐兴趣点数量	U	F	G	J
5	2.80	3.40	3.95	4.09
10	2.98	3.97	4.18	4.32
20	3.21	4.26	4.48	4.79
50	4.20	4.88	5.19	5.40

4 结束语

本文提出一种个性化联合推荐算法，通过引入兴趣点的位置因素去除不可能或可能性较小的 POI，形成初步候选集；综合考虑 POI 的类别、流行度及用户的社会行为，增加用户个性化的程度，提高推荐结果的质量。此外，通过在大规模的 Foursquare 数据集上进行了实验对比，证明了相较于其他兴趣点推荐算法，该算法的准确率和召回率都有所提高。

在将来的工作中希望能在以下两个方面取得突破：一方面，争取能将时间因素也引入推荐算法中，进一步提高推荐算法性能；另一方面，现有的兴趣点推荐算法都集中于本地推荐，希望能设计一种高性能的异地兴趣点推荐算法。

参考文献：

[1] Resnick P, Varian H R. Recommender systems [J]. Communications of the ACM, 1997, 40 (3): 56-58.

[2] Schafer J B, Konstan J A, Riedl J. E-commerce recommendation applications [J]. Data Mining & Knowledge Discovery, 2001, 5 (1-2): 115-153.

[3] Yang X, Guo Y, Liu Y, el al. A survey of collaborative filtering based social recommender systems [J]. Computer Communications, 2013, 41 (5): 1-10.

[4] Wang H, Terrovitis M, Mamoulis N. Location recommendation in location-based social networks using user check-in data [C]// Proc of the 21st ACM Sigspatial International Conference on Advances in Geographic Information Systems. 2013: 374-383.

[5] Liu B, Fu Y, Yao Z, *et al.* Learning geographical preferences for point-of-interest recommendation [C]// Proc of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2013: 1043-1051.

[6] Gao H, Tang J, Hu X. Content-aware point of interest recom-mendation on location-based social networks [C]// Proc of the 29th AAAI Conference on Artificial Intelligence. 2015: 1721-1727.

[7] Bao J, Zheng Y, Mokbel M F. Location-based and preference-aware recommendation using sparse geo-social networking data [C]// Proc of the 20th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2012: 199-208.

[8] Schafer J B, Dan F, Herlocker J, *et al.* Collaborative filtering recommender systems [M]// The Adaptive Web. Berlin: Springer, 2007: 291-324.

[9] Konstas I, Stathopoulos V, Jose J M. On social networks and collaborative recommendation [C]// Proc of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieva. 2009: 195-202.

[10] Ye M, Yin P, Lee W C, el al. Exploiting geographical influence for collaborative point-of-interest recommendation [C]// Proc of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2011: 325-334.

[11] Yuan Q, Cong G, Ma Z. Time-aware point-of-interest recom-mendation [C]// Proc of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2013: 63-372.

[12] Liu B, Fu Y, Yao Z, el al. Learning geographical preferences for point-of-interest recommendation [C]// Proc of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2013: 1043-1051.

[13] 曹玖新，董羿，杨鹏伟，等. LBSN 中基于元路径的兴趣点推荐 [J]. 计算机学报, 2016, 39 (4): 675-684.

[14] Tobler W R. A computer movie simulating urban growth in the detroit region [C]// Proc of International Geographical Union Commission on Quantitative Methods. 1970: 234-240.